Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 8, No. 4, August, 2024

Expression and Epitope Prediction of the Sirohydrochlorin Cobaltochelatase Isolated from a Local Strain of *Mycobacterium Tuberculosis*

Ahyar Ahmad [1, 2*], Abdul Karim [1], Rugaiyah A. Arfah [1], Rosana Agus [3], Rusdina Bte Ladju [4, 5], Najdah Hidayah [6], Muhammad N. Massi [7], Astutiati Nurhasanah [6], Harningsih Karim [8], Irda Handayani [9], Rizal Irfandi [10]

[1] *Department of Chemistry, Faculty of Mathematics and Natural Science, Hasanuddin University, 90245, Makassar, Indonesia.*

[2] *Research and Development Centre for Biopolymers and Bioproducts; LPPM, Hasanuddin University, Makassar-90245, Indonesia.*

[3] *Biology Department, Mathematics and Natural Science Faculty, Hasanuddin University, Makassar 90245, Indonesia.*

[4] *Department of Anatomic Pathology, Faculty of Medicine, Hasanuddin University, Makassar 90245, Indonesia.*

[5] *Hasanuddin University Medical Research Center, Faculty of Medicine, Hasanuddin University, Makassar 90245, Indonesia.*

[6] *Research Center for Vaccine and Drugs, Health Research Organization, National Research and Innovation Agency (BRIN), BJ Habibie Science and Technology Area, Serpong Center for Research, Science and Technology, Tangerang Selatan, Banten 15314, Indonesia.*

[7] *Department of Clinical Microbiology, Faculty of Medicine, Hasanuddin University, Makassar 90245, Indonesia.*

[8] *Department of Pharmacy, YAMASI School of Pharmacy, Makassar 90222, Indonesia.*

[9] *Department of Clinical Pathology, Faculty of Medicine, Hasanuddin University, Makassar 90245, Indonesia*

[10] *Department of Chemistry, Faculty of Mathematics and Natural Science, Universitas Negeri Makassar, Makassar, 90244, Indonesia.*

Abstract

The efficacy of the BCG vaccine, a widely recognized tuberculosis vaccine, has shown varying degrees of effectiveness, ranging from 0% to 80%. Sirohydrochlorin cobaltochelatase (CbiX), found in *Mycobacterium tuberculosis*, plays a crucial role in the bacteria metabolism, making it a promising target for future vaccine and drug development. Although several studies had been published regarding its role in *M. tuberculosis* vitamin B-12 metabolism, the potentials of CbiX protein as a vaccine candidate had not been widely discussed nor explored. This study focuses on the cloning and expression of the Rv0259c gene obtained from a clinical isolate of *M. tuberculosis*, as well as the exploration of CbiX protein epitopes. The Rv0259c gene was isolated by PCR, cloned into the pGEM®-TEasy vector, and subsequently sub-cloned into the pTrcHisA expression vector. Sanger sequencing, followed by BLASTN and BLASTX analyses, confirmed the presence of the CbiX protein-encoding gene. The amino acid sequence was predicted using BioEdit v.7.0.11, and a three-dimensional (3D) model was generated using SwissModel. Exploration for both B and T-cell epitopes was conducted using IEDB Ellipro, MHCI, and MHCII tools, revealing highly immunogenic epitopes, indicating the potential of CbiX as a vaccine candidate. Alignment using MAFFT between the putative amino acid sequence and CbiX proteins available in the NCBI database identified an amino acid variation (A182), situated outside the B-cell epitopes but within the T-cell epitopes. *In silico* analysis of HLA allele frequency predicted vaccine coverage of 86.14%±10.77%, with two significant epitope cores identified: AASAHPHVT and RRVAVASFL (both highly antigenic and showing high-frequency HLA allele binding), suggesting the protein might be a potential antigen for a future vaccine candidate.

Keywords:

CbiX;
Cloning;
Expression;
Mycobacterium tuberculosis;
Rv0259c;
B-cell-epitope;
T-cell-epitope.

Article History:

Received:	22	February	2024
Revised:	05	June	2024
Accepted:	17	June	2024
Published:	01	August	2024

* **CONTACT**: ahyarahmad@gmail.com

DOI: http://dx.doi.org/10.28991/ESJ-2024-08-04-07

1- Introduction

Tuberculosis (TB) remains a significant global health challenge, especially in developing nations, contributing to high mortality and morbidity rates. In 2022 alone, TB affected 7.5 million individuals worldwide, with an estimated 1.3 million fatalities [1]. Second only to HIV, TB ranks among the deadliest infectious diseases [2]. Indonesia, following India and China, faces a substantial burden of TB cases, recording 16,528 deaths out of 724,309 cases in 2021 [3]. Control and prevention efforts are hindered by the lack of effective treatments and precise diagnostic methods. Multidrug-resistant tuberculosis (MDR-TB) compounds the challenge due to *Mycobacterium tuberculosis*' unique characteristics, including slow growth, diverse strains, and increased virulence, leading to latent infections in 45–55% of cases [4]. Moreover, existing TB vaccines like BCG are losing effectiveness, while diagnostic methods remain subpar.

Although the BCG vaccine has been instrumental in TB prevention programs [5], its efficacy against MDR-TB and recent TB outbreaks is limited. While it protects infants from certain TB forms, it offers inadequate protection against pulmonary TB in adults, particularly those with latent TB reactivation [6]. In addition, its effect diminishes over time, with protection limited to between 10 – 20 years [7]. Addressing current TB management challenges requires improved diagnostic techniques and exploration of potential vaccine candidates.

In general, the development of novel TB vaccines focuses on several different objectives. According to their purposes, these vaccines are classified as follows: priming vaccines, designed to offer superior and longer-lasting immunoprotective effects compared to the existing BCG vaccine; booster vaccines, serving as supplementary vaccines to enhance or prolong and improve BCG protectiveness; latent infection preventive vaccines, aimed at preventing the endogenous reactivation of MTB in infected individuals while guarding against exogenous reinfection; and therapeutic vaccines, serving as effective complements to conventional chemotherapy. These therapeutic vaccines aim to reduce lesions, facilitate sputum negativity for MTB, promote closure of cavities, shorten treatment duration, reinforce the effects of chemotherapy, and promote the outcome of the disease, effectively treating drug-resistant TB [7].

A significant challenge in TB vaccine development stems from limited knowledge of protective antigens and the absence of reliable preclinical evaluation indicators [7]. *M. tuberculosis* has thousands of genes encoding thousands of proteins; however, which of these proteins would be the best antigens and how many proteins should be included in a vaccine is yet to be revealed. Moreover, the complexity of TB infection necessitates a multi-antigen approach to enhance vaccine efficacy across diverse populations. Integrating immune-dominant antigens from different Mtb sites into a vaccine represents a promising strategy to bolster its immunogenicity and potency against TB [8].

Amongst potential TB vaccine candidates, those that reach clinical assessments include viral vectored, live-attenuated, cellular extract, and protein subunit vaccines [7]. As subunit vaccines contain only a small number of selected antigens, a major challenge in protein subunit vaccine development is the need to determine the optimal antigen combination in the vaccine that induces immune diversity. The current advancement of bioinformatics and structural biology techniques, however, has opened the door to the possibility of using novel approaches such as *in silico* screening, target identification, exploration, and optimization of protective antigens [9, 10], hence having tremendous potential in the development of TB vaccines.

Sirohydrochlorin cobaltochelatase (CbiX) is a protein found in various *Mycobacterium* species, including *M. tuberculosis*, *M. abscessus*, and *M. gastri*. Encoded by the Rv0259c gene in *M. tuberculosis* [11], recent annotations suggest it is a putative lyase [12] involved in the cobalamin (vitamin B12) biosynthesis [11, 13]. Cobalamin influences cell metabolism through enzyme cofactor activity and gene expression regulation via riboswitches. *M. tuberculosis* harbors two vitamin B12-dependent riboswitches in its genome. The first one is encoded by Rv1133c, regulating the *metE* gene expression of the cobalamin-independent methionine synthase, and the other one is encoded by Rv0256c, affecting the PPE2-cobQ1-cobU operon genes, which is presumably involved in vitamin B12 synthesis [11]. While researchers explore riboswitches as potential drug targets [14], their suitability as targets for subunit protein vaccines remains unexplored. Despite numerous studies on its role in *M. tuberculosis* vitamin B12 metabolism, the potential of the CbiX protein as a vaccine candidate has not been extensively discussed or explored.

This study aims to clone and express *M. tuberculosis* Rv0259c genes isolated from clinical samples in Makassar, Indonesia, and predict the CbiX protein-specific B and T-cell epitopes. The process involved cloning the CbiX-encoding Rv0259c gene into the pGEM®-TEasy vector, resulting in the formation of the pGEM®-TEasy-Rv0259c recombinant plasmid. Subsequently, the Rv0259c gene was subcloned from this recombinant plasmid into the pTrcHisA expression vector at the *Bam*HI/*Hin*dIII cloning site, generating the pTrcHisA-Rv0259c plasmid. Expression of the recombinant His-tagged CbiX protein was achieved in *E. coli* BL21 cells. Epitope exploration was conducted *in silico* using the three-dimensional (3D) model of the CbiX protein, constructed based on the nucleotide sequence of the isolated Rv0259c gene. The antigenicity and affinity to HLA of the potential epitopes were analyzed to assess the protein's potential as a vaccine candidate.

2- Experimental Section

2-1- Mycobacterium Tuberculosis

M. tuberculosis isolate utilized in this study originated from the culture collection of Wahidin Sudirohusodo Hospital in Makassar, Indonesia [15]. All procedures involving human subjects adhered to ethical standards, with signed, written informed consent obtained in accordance with the Ethics Committee at Hasanuddin University Hospital, Makassar, Indonesia. The culture has been consistently maintained through sub-culturing or refreshing every 2-3 weeks to ensure viability and to obtain isolates in the logarithmic phase (exponential growth phase).

2-2- Amplification of Rv0259c Gene Encoding CbiX Protein

Primers were designed using the Primer-BLAST tool* [16], based on the nucleotide sequence of *M. tuberculosis* H37Rv (ATCC 27294) genome retrieved from the GenBank database (accession number NC_000962.3) The primer sequences are as follows; forward: 5'-GCGGGATCCATGAACCTGATCTTGACGGC-3'; and reverse: 5'-CCGAAGCTTTCAAAGATCAAGCGTCACCGG-3'. These primers were constructed to include the *Bam*HI restriction enzyme site upstream from the start codon ATG and the *Hind*III site downstream from the stop codon TGA.

The Rv0259c gene was amplified using the PCR, as detailed in previous studies [15, 17, 18]. The total volume of the PCR reaction was 30 µL, comprising 15 µL PCR mix, forward (2 µL) and reverse (2 µL) primers at a concentration of 10 pmol/µL, 9 µL of ddH_2O, and 2 µL of *M. tuberculosis* genomic DNA from the Wahidin Sudirohusodo Hospital collection used as the template. The PCR process included an initial denaturation phase at 95°C for 5 minutes, then 30 cycles of denaturation at 95°C for 2 minutes, annealing at 55°C for 30 seconds, and extension at 72°C for 1 minute. A 10-minute hold at 72°C marked the end of the process. Subsequently, the PCR tubes were cooled for 10 minutes at 4°C. The PCR products were electrophoresed on a 1.5% agarose gel in 1X TAE buffer, and the results were observed and documented using a UV transilluminator [19].

2-3- Purification of Rv0259c Amplicon

PCR product purification was conducted using an ez-10 column and a Geneaid kit (Qiagen) according to the manufacturer's instructions. Electrophoresis on a 1.5% agarose gel verified the PCR product purification results [19]. The size of the Rv0259c gene (Accession Number: NC_000962.3, Gene ID: 886657), available in the GenBank †, was used the standard (744 bp).

2-4- Rv0259c Gene Ligation into pGEM®-TEasy Cloning Vector and Transformation to E. coli JM109

After the PCR product was purified, it was combined with the pGEM®-TEasy vector (Promega). The mixture was incubated with T4 DNA ligase at 15 °C for at least four hours or at 4 °C overnight. Subsequently, the ligation mixture was transformed into *E. coli* cells (JM109 strain) using the heat shock method at 42 °C for 30 seconds. The transformation products were spread on LB/ampicillin/IPTG/X-Gal agar medium as previously described [15].

2-5- Screening of E. coli JM109 Carrying pGEM®-TEasy-Rv0259c

Blue-white colony screening was performed to obtain positive transformants [20]. *E. coli* JM109 carrying pGEM®-TEasy with inserts typically appeared white [20]. The white colonies were screened further by colony PCR, using the same primers and methods (except for the pre-denaturation step, which was extended by 5 minutes to ensure bacterial cell lysis) as those used to isolate the Rv0259c gene from the *M. tuberculosis* genome. Blue colonies were used as a negative control, whereas genomic DNA of *M. tuberculosis* H37Rv was used as a positive control. The size of the Rv0259c gene (Accession Number: NC_000962.3, Gene ID: 886657), available in the GenBank †, served as a standard (744 bp). Colonies carrying recombinant plasmid with the correct insert size, as indicated by the correct colony PCR amplicon size, were selected, and the insert was confirmed by Sanger sequencing. BioEdit v.7.0.11 [21], NCBI BLASTN [22], and BLASTX [23] were used to analyze the nucleotide sequence.

2-6- Generation of the pTrcHisA-Rv0259c Plasmid

Plasmids were isolated from three white *E. coli* JM109 colonies, previously confirmed to carry pGEM®-TEasy-Rv0259c. The pGEM®-TEasy-Rv0259c plasmid underwent digestion with a combination of *BamH*I/*Hind*III restriction enzymes. The restriction fragment containing Rv0259c was extracted from agarose gel, followed by ligation to the pTrcHisA expression vector, which had been pre-digested with the same restriction enzymes. Ligation was catalyzed by T4 DNA ligase at 15 °C for a minimum of four hours or overnight at 4 °C. The resulting ligation product was used to transform competent *E. coli* BL21 (DE3) cells. Transformants were screened on LB+Ampicillin agar medium. The cloning procedures involving the pGEM®-TEasy-Rv0259c cloning vector and the pTrcHisA-Rv0259c expression vector are presented in Figure 1.

* *https://www.ncbi.nlm.nih.gov/tools/primer-blast/*
† *https://www.ncbi.nlm.nih.gov/gene/886657*

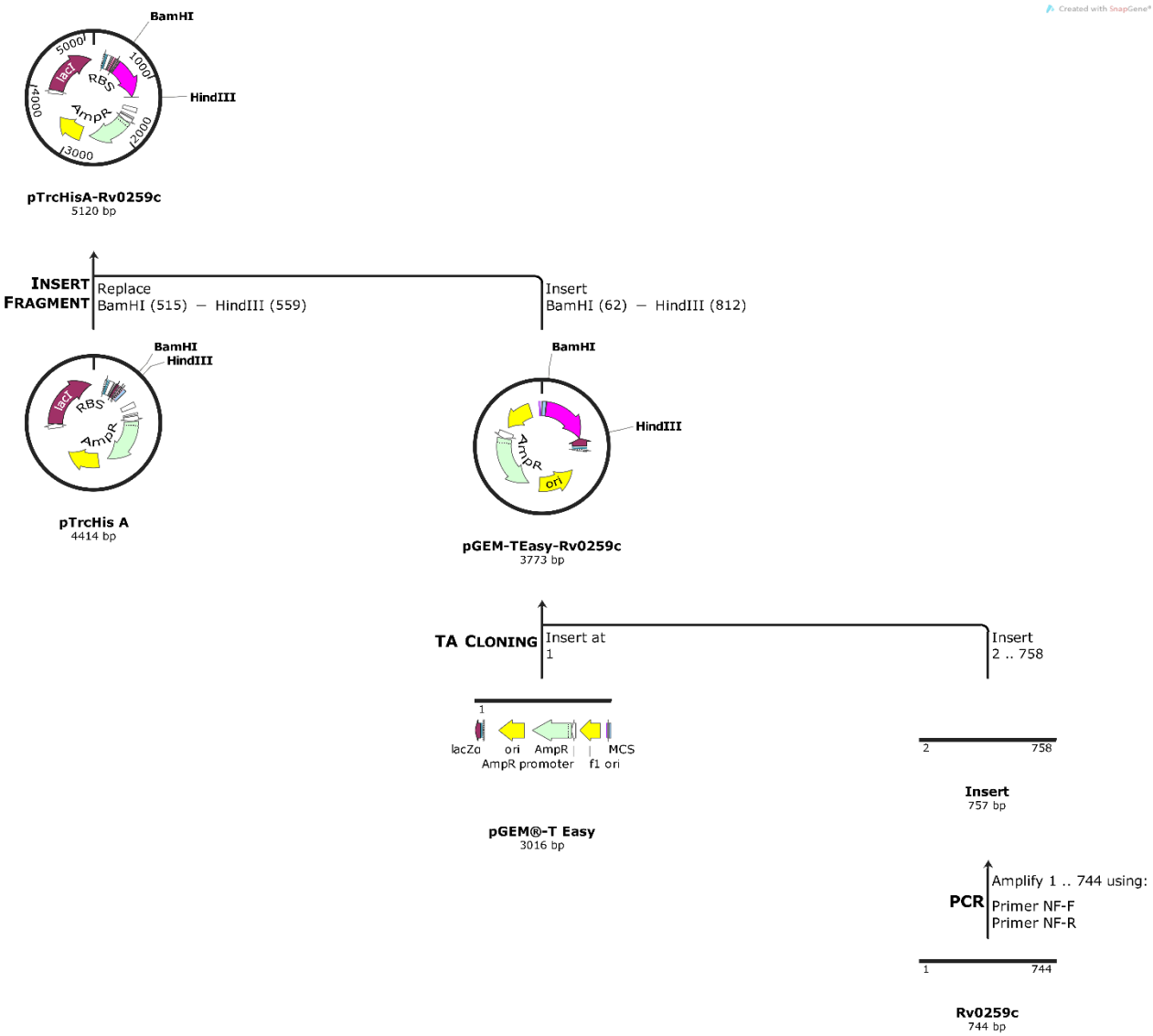

Figure 1. Isolation and cloning procedures of the Rv0259c gene from the M. tuberculosis genome (diagram generated using SnapGene®)

2-7- Three-Dimensional (3D) Prediction of CbiX Recombinant Protein

Prediction of the recombinant protein 3D structure was carried out using SwissModel* [24], employing the crystal structure of *Bacillus subtilis* sirohydrochlorin ferrochelatase (SirB protein) (PDB ID: 5ZT8.1)† [25] from the RCSB Protein Data Base‡ [26] as a template.

2-8- Prediction of Immunogenic Properties of CbiX Recombinant Protein

B and T cell epitope analysis was conducted using the IEDB database. Exploration of the B cell epitopes followed the previously described method [27], employing the IEDB Ellipro Antibody Epitope Prediction tool§ [28]. Only peptides with a score > 0.5 were considered as B-cell epitopes. Continuous and discontinuous epitopes were predicted.

T-cell epitopes were predicted using IEDB MHCI Binding Prediction** and IEDB MHCII Binding Prediction†† tools. HLA-I binding sites were predicted using the NetMHCPan BA method [29-31], whereas HLA-II binding sites were predicted using the NN-align method [32, 33]. Predicted sites were selected based on the rank and IC50 values, with only those having rank < 1 and IC50 < 50 being considered for further antigenicity analysis in VaxiJen v.2.0‡‡ [34]. Only peptides with VaxiJen antigenicity score > 0.4 were identified as potential T cell epitopes.

** https://swissmodel.expasy.org/*

† https://doi.org/10.2210/pdb5ZT8/pdb

‡ http://www.rcsb.org/

§ http://tools.iedb.org/ellipro/

*** http://tools.immuneepitope.org/mhci/*

†† http://tools.iedb.org/mhcii/

‡‡ https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html

2-9- Prediction of Epitope Population Coverage

The population coverage of the T-cell epitopes was predicted using the IEDB Population Coverage tool* [35], as previously published [36]. Analysis was performed in combination between HLA-I and HLA-II, on all antigenic core peptides (26 core peptides) uncovered by the T–cell epitope exploration, across various regions available in the tool (i.e., East Asia, Northeast Asia, South Asia, Southeast Asia, Southwest Asia, Europe, East Africa, West Africa, Central Africa, North Africa, South Africa, West Indies, North America, Central America, South America, and Oceania). Models displaying the positions of selected core peptides were generated in Protean3D®.

3- Results

3-1- Amplification and Isolation of Rv0259c Gene

The Rv0259c gene was amplified using the PCR following procedures outlined in previous studies [15, 17, 18], resulting in distinct DNA bands with an estimated size of 744 bps, as illustrated in Figure 2-A. No DNA band was shown in the negative PCR control (lane 1). The size of the PCR product corresponded to that of the Rv0259c gene (Accession Number: NC_000962.3, Gene ID: 886657), available in the GenBank†. Subsequently, the purified PCR product (Figure 2-B) was ligated into pGEM®-Teasy.

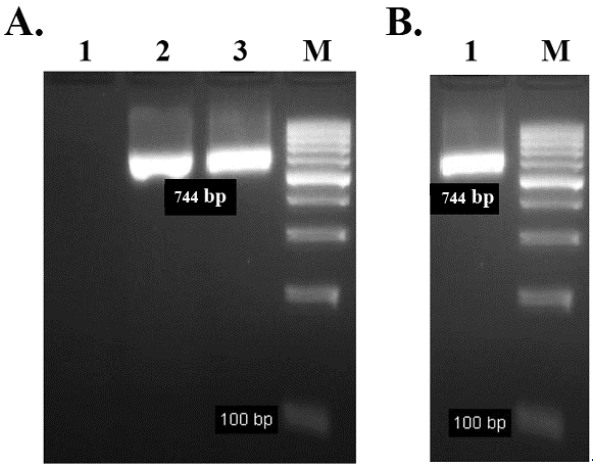

Figure 2. Visualization on 1.5% agarose gel. A. PCR product of Rv2059c gene and B. purified Rv0259c gene PCR product

3-2- Cloning of Rv0259c Gene into pGEM®-TEasy Cloning Vector and Transformation to E. coli JM109

E. coli JM109 transformants carrying pGEM®-TEasy-Rv0259c were screened on LB agar medium supplemented with ampicillin, X-gal, and IPTG (LB/Amp/IPTG/X-gal medium) (Figure 3). Absence of colonies on the negative control plate (Figure 3-A) contrasted with the flourishing growth of the positive control (*E. coli* JM109 carrying pGEM®-TEasy plasmid) (Figure 3-B), indicating successful screening. Evaluation of the transformant plate (Figure 3-C) revealed 347 colonies, comprising 158 blue colonies and 189 white colonies, indicating 54.5% positive transformant rate.

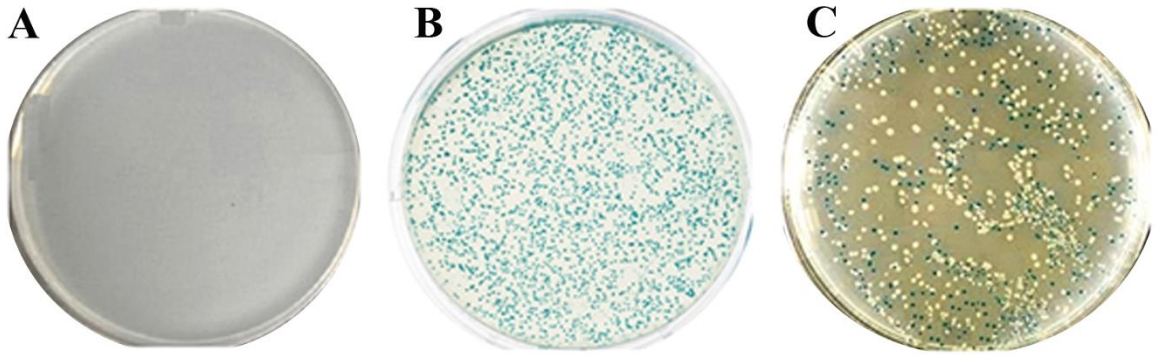

Figure 3. Screening of E. coli JM109 recombinant was performed on LB/Amp/IPTG/Xgal agar medium. LB/Amp/IPTG/Xgal agar plates were employed for the cultivation of A. negative control (E. coli JM109 without plasmid), B. positive control (E. coli JM109 cells with pGEM®-TEasy control plasmid), and C. E. coli JM109 cells transformed with the ligation product. White colonies from plate C were subsequently chosen and subjected to further screening through colony PCR.

* *http://tools.iedb.org/population/*

† *https://www.ncbi.nlm.nih.gov/gene/886657*

Colony PCR results of selected white colonies were visualized on 1.5% agarose gel electrophoresis, demonstrating the presence of a single DNA band measuring 744 bps (Figure 4-A). This observation indicated that the plasmid DNA within the selected white colonies carried an insert of the same size as the Rv0259c gene.

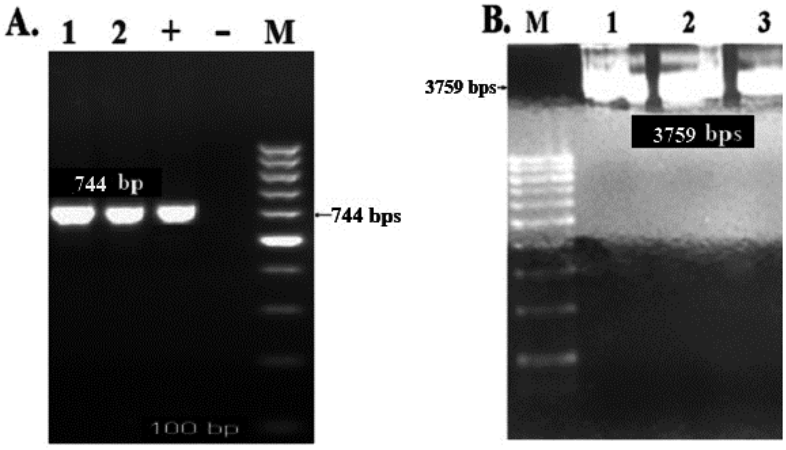

Figure 4. A. Visualisation of PCR colony results of selected white colonies (Lane 1 – 2), a blue colony used as negative control (Lane -), and M. tuberculosis H37Rv genomic DNA used as positive control (Lane +). B. Recombinant plasmid isolated from three selected white colonies.

The recombinant plasmid exhibiting the correct colony PCR pattern was sent for Sanger sequencing. The nucleotide sequence was analyzed using BioEdit v.7.0.11 [21], NCBI BLASTN [22], and BLASTX [23]. NCBI BLASTN showed 100% identity (E value 0.0, and Query Cover 100%) to Rv0259c (MTCY06A4.03c) of the *M. tuberculosis* H37Rv genome, whilst BLASTX predicted 100% identity to sirohydrochlorin chelatase protein of *M. tuberculosis* H37Rv (accession number: WP_003916613.1), with Query Cover 99% and E value $1x10^{-148}$. These findings confirmed the successful isolation and cloning of the Rv0259c gene, making it ready for insertion into the expression vector for subsequent steps in the experimental process.

3-3- Subcloning of Rv0259c Gene into Expression Vector pTrcHisA

The isolated pGEM®-TEasy-Rv0259c plasmid (Figure 4-B) was double digested with *Bam*HI/*Hin*dIII restriction enzymes prior to ligation to pTrcHisA plasmid, which was similarly digested using the same enzymes. The restriction digest (*Bam*HI/*Hin*dIII) result of the plasmid isolated from one of the selected white colonies was visualized on 1.5% agarose gel electrophoresis (Figure 5-D Lane 2), alongside a 100bp DNA Ladder (Invitrogen) as a molecular weight marker (Figure 5-D). For comparison, linear pGEM®-TEasy was also electrophoresed (Figure 5-D Lane 1). Two fragments were generated from the restriction digest of pGEM®-TEasy-Rv0259c, 3023 and 744 bp, the longer fragment corresponds to the plasmid backbone, whilst the shorter one corresponds to the gene. This outcome is consistent with the simulation performed in SnapGene® (Figure 5-C)

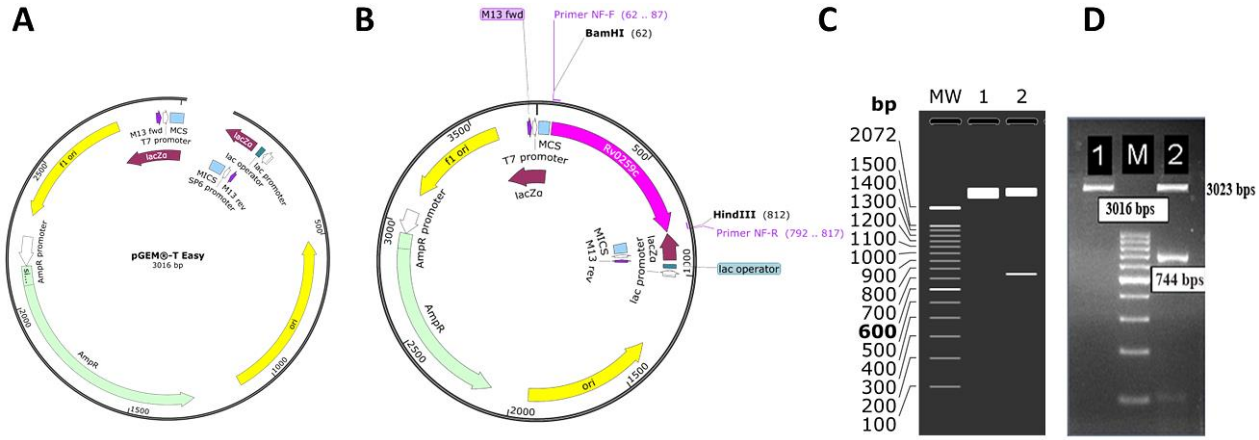

Figure 5. Map of pGEM®-TEasy (A) and pGEM®-TEasy carrying Rv2059c gene (B). SnapGene® agarose gel simulation of linear pGEM®-TEasy without an insert (3016 bp) (Lane 1), and BamHI/HindIII restriction digest of the pGEM®-TEasy-Rv0259c (Lane 2) (C). Visualization of the restriction digest result of the vector containing the insert through agarose gel electrophoresis (D). Lane 1 shows linear pGEM®-TEasy without an insert, whereas Lane 2 displays BamHI/HindIII restriction digest result of the pGEM®-TEasy-Rv0259c. DNA marker (M): 100 bp DNA Ladder (Invitrogen).

The gel-purified restriction fragment containing the Rv0259c gene was utilized for ligation to the previously double-digested pTrcHisA plasmid (*Bam*HI/*Hin*dIII). The resulting ligation product (Figure 6-D) was used to transform *E. coli* BL21 (DE3), which underwent screening in LB/Amp medium to obtain positive transformant (Figure 6-B). Subsequently, the positive transformants were further screened by their capability to express CbiX protein (Figure 6-C).

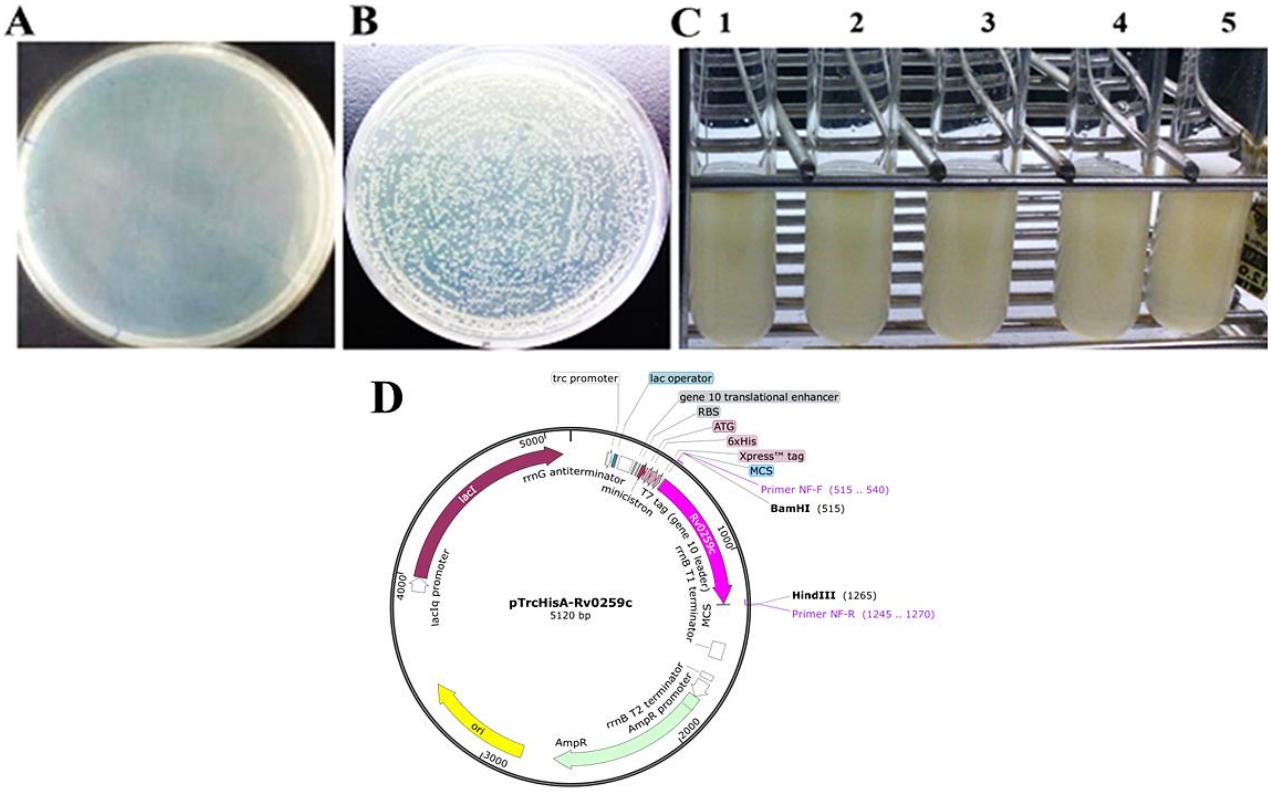

Figure 6. The growth of E. coli BL21(DE3) without (A) and with (B) the recombinant plasmid pTrcHisA-Rv0259c on LB agar plates. Liquid culture (C) in different tubes containing LB liquid medium as control (tube 1), culture of *E. coli* BL21(DE3) pTrcHisA-Rv0259c without IPTG (tubes 2 and 3), and induced with 30 µM IPTG (tubes 4 and 5). Map of pTrcHisA-Rv0259c recombinant expression plasmid (D).

3-4- Expressed Protein Characterization of the SDS-PAGE Method

The His-tagged CbiX proteins were purified by affinity chromatography utilizing the Ni-NTA Superflow Agarose matrix. As depicted in Figure 8, the pellet obtained from the recombinant culture induced with 30 µM IPTG for four hours exhibited a distinct protein band that appeared unique, blue in color, and noticeably thicker. The estimated molecular weight of the CbiX recombinant protein band was approximately 26 kDa. Conversely, in the absence of inducers (IPTG), the culture failed to yield a detectable protein band with the correct molecular weight (data not shown), confirming that the 26 kDa band was indeed induced by IPTG. Given that the protein was primarily detected in the lysate supernatant (Figure 8 Lane 5), it can be inferred that the CbiX protein is expressed intracellularly as soluble proteins within the cells.

3-5- Analysis of Rv0259c Sequence and Generation of 3D Protein Model

The amino acid sequence of the recombinant protein, predicted using BioEdit v.7.0.11 [21], is displayed in Figure 7. The Rv0259c gene spans 744 bp, commencing with the start codon ATG (Met) and concluding with TGA stop codon, thus encoding for a protein of 247 amino acids.

```
>Rv0259c

1     ATG AAC CTG ATC TTG ACG GCC CAC GGA ACC CGC AGA CCG TCC GGT   45
1     Met Asn Leu Ile Leu Thr Ala His Gly Thr Arg Arg Pro Ser Gly   15

46    GTC GCG ATG ATA GCG GAC ATC GCC GCG CAA GTG AGC GCG CTC GTC   90
16    Val Ala Met Ile Ala Asp Ile Ala Ala Gln Val Ser Ala Leu Val   30

91    GAC CGC ACT GTG CAG GTC GCC TTC GTT GAT GTG CTC GGA CCC TCA   135
31    Asp Arg Thr Val Gln Val Ala Phe Val Asp Val Leu Gly Pro Ser   45

136   CCC AGC GAA GTG CTT TCC GCC CTT TCC TGC CGT CCC GCA ATC GTG   180
46    Pro Ser Glu Val Leu Ser Ala Leu Ser Cys Arg Pro Ala Ile Val   60

181   GTG CCC GCG TTC TTG TCC CGC GGA TAT CAT GTT CGC ACC GAC CTG   225
61    Val Pro Ala Phe Leu Ser Arg Gly Tyr His Val Arg Thr Asp Leu   75

226   CCC GCT CAT GTC GCA GCC AGT GCG CAC CCG CAT GTC ACG GTC ACC   270
76    Pro Ala His Val Ala Ala Ser Ala His Pro His Val Thr Val Thr   90

271   CCC GCG TTG GGG CCA TGC CGT GAG ATC GCG CAG ATA GTC ACC CAG   315
91    Pro Ala Leu Gly Pro Cys Arg Glu Ile Ala Gln Ile Val Thr Gln   105

316   CAG CTG GTG GAA TCC GGT TGG CGT CCT GGC GAT TCA GTG ATC CTC   360
106   Gln Leu Val Glu Ser Gly Trp Arg Pro Gly Asp Ser Val Ile Leu   120

361   GCA GCG GCC GGC GCG TCG GAT AGG CGG GCA CGT GCC GAC CTG CAC   405
121   Ala Ala Ala Gly Ala Ser Asp Arg Arg Ala Arg Ala Asp Leu His   135

406   ACC ACT CGG ACG CTA GTG TCC GAA CTG ACC GGA TCA TGG GTG GAC   450
136   Thr Thr Arg Thr Leu Val Ser Glu Leu Thr Gly Ser Trp Val Asp   150

451   ATG GGA TTT GCG GGC ACC GGC GGT CCG GAT GTG CGC ACG GCT GTG   495
151   Met Gly Phe Ala Gly Thr Gly Gly Pro Asp Val Arg Thr Ala Val   165

496   CAG CGG GCC CGA GAC CGG GCC GAA GCC AAC CGA GGT GCG CGC CGC   540
166   Gln Arg Ala Arg Asp Arg Ala Glu Ala Asn Arg Gly Ala Arg Arg   180

541   GTC GCG GTC GCC TCC TTC CTG TTG GCA GAA GGC CTT TTT CAA GAG   585
181   Val Ala Val Ala Ser Phe Leu Leu Ala Glu Gly Leu Phe Gln Glu   195

586   CGG CTG CGG GCA TCC GGT GCC GAT GTG GTA ACC CGA CCG CTC GGC   630
196   Arg Leu Arg Ala Ser Gly Ala Asp Val Val Thr Arg Pro Leu Gly   210

631   ACC CAT CCG GGC TTG GCA CAG CTG GTC GCG AAC CGA TTT CGC AGC   675
211   Thr His Pro Gly Leu Ala Gln Leu Val Ala Asn Arg Phe Arg Ser   225

676   GCG GTA GCA CGT CAG CAA CGG CTG CAC CGC TGG CAT GGC ACG CCG   720
226   Ala Val Ala Arg Gln Gln Arg Leu His Arg Trp His Gly Thr Pro   240

721   ACA CCG GTG ACG CTT GAT CTT TGA   744
241   Thr Pro Val Thr Leu Asp Leu End
```

Figure 7. Predicted CbiX protein amino acid sequence from isolated Rv0259c gene sequencing results

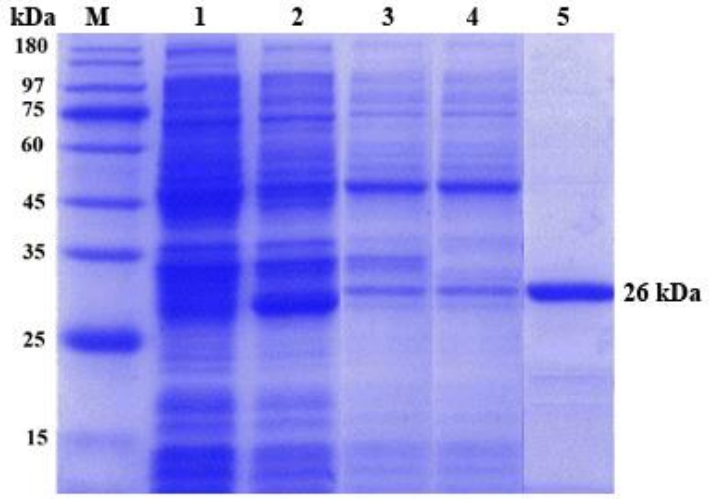

Figure 8. Purification of the recombinant CbiX protein. Molecular weight marker (M), induced whole cell lysates (Lane 1), wash flowthrough (Lane 2), flowthrough of elution with 0.5 M Tris-Cl buffer with 100 mM Imidazole pH 7.3 (Lane 3), 200 mM Imidazole pH 5.9 (Lane 4), and 300 mM Imidazole pH 4.5 (Lane 5), CbiX protein bands at 26 kDa are shown.

The putative 3D structure of the CbiX protein, constructed based on the predicted amino acid sequence in SwissModel [24] (*accessed 1 March 2023*), is presented in Figure 9-A. The model was generated using template *B. subtilis*

sirohydrochlorin ferrochelatase (SirB protein) (PDB ID: 5ZT8.1)* [25], which shares 18.50% sequence identity to the CbiX protein and shows Global Model Quality Estimate (GMQE) score of 0.57. Amino acids 2 – 233 of CbiX were used in the development of the model (94% coverage). 3D alignment between the CbiX model and the SirB protein template is presented in Figure 9-B, while 2D alignment between CbiX and 5ZT8.1 is presented in Figure 9-C, where the residue colouring indicates the confidence score of the model. The prediction of local similarity between the CbiX protein and the template (QMEAN) is displayed in Figure 9-D, and the average local model quality estimation score (global QMEANDisCo score) was 0.62±0.06. QMEAN is a measure of model accuracy, which is developed according to statistical potentials of mean force to generate global and per residue quality estimates. The QMEAN value ranges from 0 – 1, with higher numbers indicating higher expected quality. This CbiX protein model was subsequently used to predict the positions of B cell epitopes

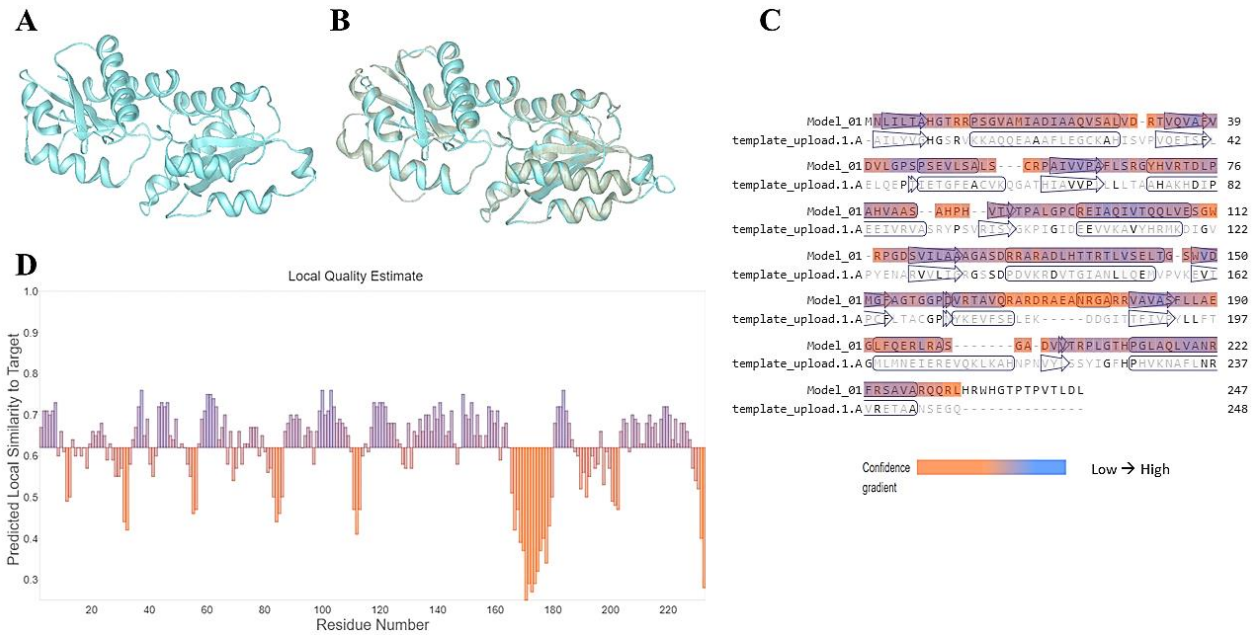

Figure 9. Predicted 3D structure of the expressed CbiX protein (A), and its (cyan) comparison with the template protein 5ZT8.1 (beige) (B). The model's Global Model Quality Estimate (GMQE) score was 0.57, and the QMEANDisCo score was 0.62±0.06. Secondary structure alignment of CbiX against template 5ZT8.1 (C). Local quality estimate of CbiX 3D model, based on similarity between CbiX and the template protein (5ZT8.1) (D).

3-6- Epitope Exploration and Prediction of Immunogenic Properties of CbiX Recombinant Protein

Based on the putative 3D structure illustrated in Figure 9, Ellipro analysis predicted the presence of ten B-cell linear epitopes, with scores ranging from 0.608 to 0.818 (refer to Table 1). The highest immunogenicity score was attributed to amino acid residues 226 to 235 (a sequence of 10 amino acid residues) featuring the sequence VARQQRLHRW. The locations of these linear epitopes are presented in Figure 10.

Table 1. Predicted B – cell linear epitopes present in CbiX protein

No	Start	End	Peptide	Number of residues	Score
1	226	235	VARQQRLHRW	10	0.818
2	50	56	SALSCRP	7	0.785
3	154	179	GTGGPDVRTAVQRARDRAEANRGARR	26	0.757
4	107	115	VESGWRPGD	9	0.751
5	76	86	AHVAASAHPHV	11	0.73
6	9	16	TRRPSGVA	8	0.696
7	23	34	AQVSALVDRTVQ	12	0.674
8	194	203	ERLRASGADV	10	0.656
9	124	127	ASDR	4	0.645
10	38	47	VDVLGPSPSE	10	0.608

* *https://doi.org/10.2210/pdb5ZT8/pdb*

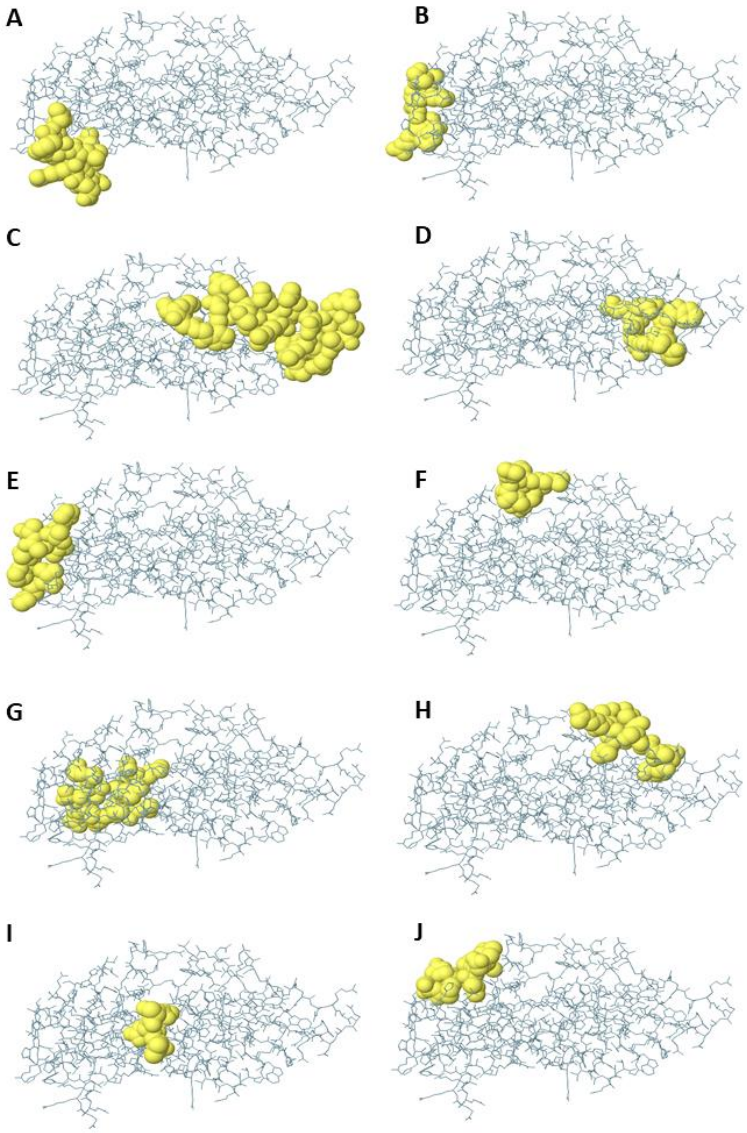

Figure 10. Positions of the predicted B – cell linear epitopes, represented by yellow balls, according to Ellipro, where A – J are epitopes 1 – 10, respectively

Furthermore, Ellipro analysis identified eight predicted discontinuous B-cell epitopes (Table 2), the positions of which are presented in Figure 11. The most highly antigenic discontinuous epitope (score 0.835) was formed by the interaction of five amino acids (V226, A227, R228, Q229, Q230).

Table 2. Predicted B – cell discontinuous epitopes of CbiX protein

No.	Residues	Number of residues	Score
1	A:V226, A:A227, A:R228, A:Q229, A:Q230	5	0.835
2	A:A23, A:Q24, A:S26, A:A27, A:L28, A:V29, A:D30, A:R31	8	0.751
3	A:S109, A:G110, A:W111, A:R112, A:P113, A:G114, A:D115, A:G123, A:A124, A:S125, A:D126, A:T138, A:S141, A:E142, A:L143, A:T144, A:G145, A:S146, A:W147, A:A153, A:G154, A:T155, A:G156, A:G157, A:P158, A:D159, A:R161, A:T162, A:A163, A:V164, A:Q165, A:R166, A:A167, A:R168, A:D169, A:R170, A:A171, A:E172, A:A173, A:N174, A:R175, A:G176, A:A177, A:R178, A:R179	45	0.719
4	A:G8, A:T9, A:R10, A:R11, A:P12, A:S13, A:V15, A:A16, A:D20, A:T32, A:V33, A:Q34, A:V38, A:D39, A:V40, A:L41, A:G42, A:P43, A:S44, A:S46, A:E47, A:S50, A:A51, A:L52, A:S53, A:C54, A:R55, A:P56, A:A76, A:H77, A:A79, A:A80, A:S81, A:A82, A:H83, A:P84, A:H85, A:V86, A:R234	39	0.696
5	A:E194, A:R197, A:A198, A:S199, A:G200, A:A201, A:D202, A:V203	8	0.689
6	A:R71, A:T72, A:D73	3	0.618
7	A:E189, A:G190, A:L191, A:R195	4	0.607
8	A:Q104, A:V107, A:E108, A:R206	4	0.561

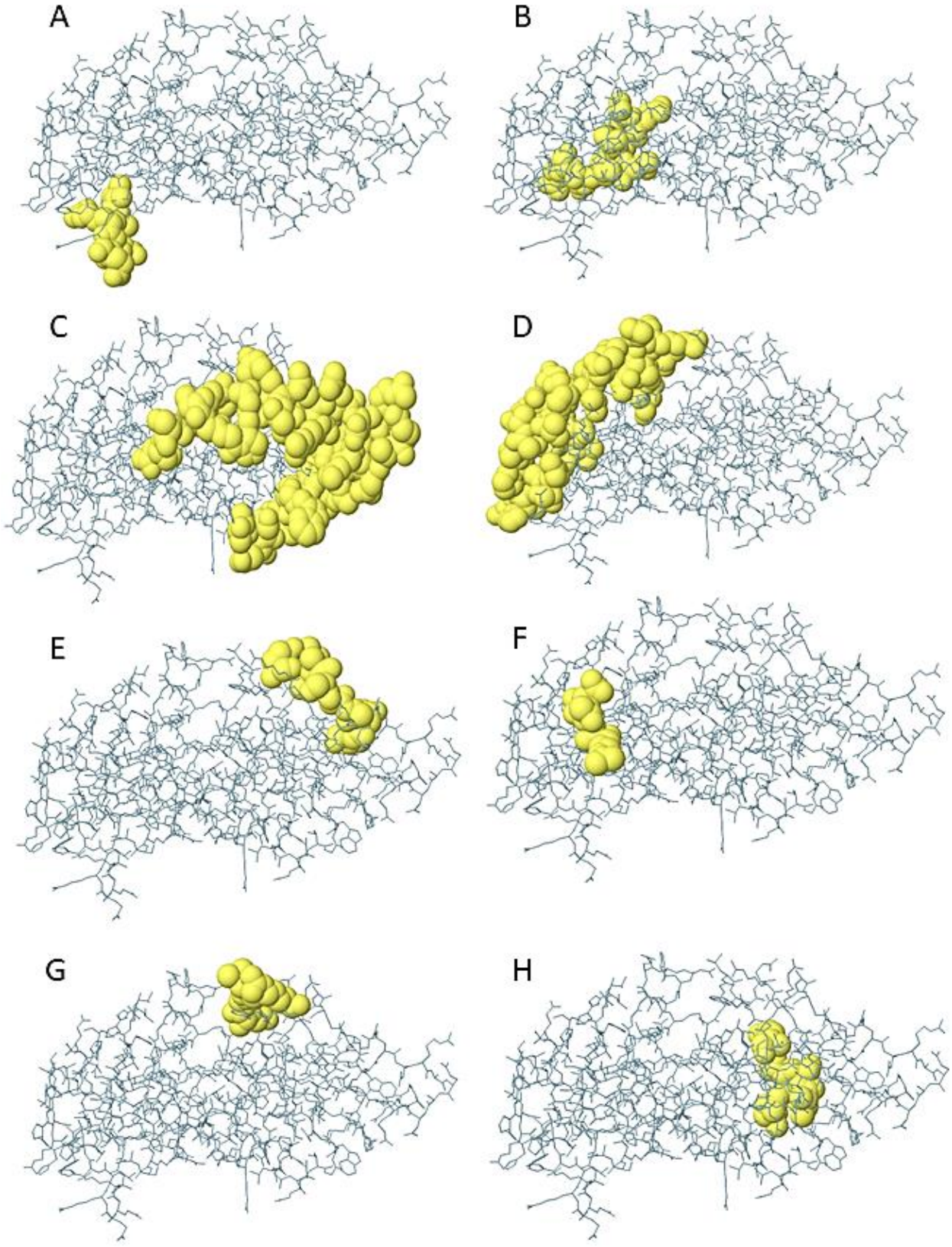

Figure 11. Positions of the predicted B–cell discontinuous epitopes, represented by yellow balls, according to Ellipro, where A – H are epitopes 1 – 8, respectively

Analysis of HLA-I binding [29–31] revealed four peptide sequences, which are highly immunogenic (VaxiJen score > 1, shaded in grey), amongst 18 predicted antigenic peptides that bind HLA-I. These peptides and their HLA-I binding sites are presented in Table 3. One of these highly antigenic peptides (RPAIVVPAF) binds to multiple HLA-I types (HLA-B*07:02 and HLA-B*35:01).

Table 3. Antigenic peptides that are predicted to bind HLA-I (VaxiJen score > 0.4). Highly antigenic peptides (VaxiJen score > 1) are shaded in grey

No.	Core	Peptide	Allele	VaxiJen Score	Amino acid position	
					Start	End
1	AFLSRGYHR	AFLSRGYHVR	HLA-A*31:01	0.7375	63	72
2	ALSCRPAIV	ALSCRPAIV	HLA-A*02:03	1.1172	52	60
3	AMADIAAQV	AMIADIAAQV	HLA-A*02:03	0.4037	17	26
	AMADIAAQV	AMIADIAAQV	HLA-A*02:01	0.4037	17	26
4	AIADIAAQV	AMIADIAAQV	HLA-A*02:06	0.4037	17	26
	AIADIAAQV	AMIADIAAQV	HLA-A*68:02	0.4037	17	26
5	AQLVANRFR	AQLVANRFR	HLA-A*31:01	0.4945	216	224
6	EVLSALSCR	EVLSALSCR	HLA-A*68:01	0.6812	48	56
	EVLSALSCR	EVLSALSCR	HLA-A*33:01	0.6812	48	56
7	GTRRPSGVA	GTRRPSGVA	HLA-A*30:01	0.8126	9	17
8	HPHTVTPAL	HPHVTVTPAL	HLA-B*07:02	1.3093	84	93
9	HVAASAHHV	HVAASAHPHV	HLA-A*68:02	0.9221	78	87
10	IVVPAFLSR	IVVPAFLSR	HLA-A*68:01	0.7489	59	67
11	RARDLHTTR	RARADLHTTR	HLA-A*31:01	0.9719	129	138
12	RFRSAVQQR	RFRSAVARQQR	HLA-A*31:01	0.5374	222	232
13	RPAIVVPAF	RPAIVVPAF	HLA-B*07:02	1.3726	56	64
	RPAIVVPAF	RPAIVVPAF	HLA-B*35:01	1.3726	56	64
14	RPAIVVPAL	RPAIVVPAFL	HLA-B*07:02	1.4766	56	65
15	RSAVARQQR	RSAVARQQR	HLA-A*31:01	0.4536	224	232
16	RTAVQRARR	RTAVQRARDR	HLA-A*31:01	0.5438	162	171
17	RVAVASFLL	RVAVASFLL	HLA-A*32:01	0.9209	180	188
18	TVTPAGPCR	TVTPALGPCR	HLA-A*68:01	0.663	88	97
19	VLGPSPSEV	VLGPSPSEV	HLA-A*02:03	0.7381	41	49
	VLGPSPSEV	VLGPSPSEV	HLA-A*02:01	0.7381	41	49

Exploration of HLA-II epitopes [32, 33] identified seven core peptides despite the varying lengths of the peptides recognized as HLA-II binders (Table 4). Amongst these peptides, 17 exhibited high antigenicity score (VaxiJen score > 1, shaded in grey). Notably, one core peptide (RVAVASFLL) appears to bind both HLA-I and HLA-II, and is recognised by HLA-A*32:01 (HLA-I) and HLA-DPA1*02:01/DPB1*14:01 (HLA-II).

Table 4. HLA-II antigenic epitopes (VaxiJen score > 0.4) of CbiX protein revealed from IEDB MHCII analysis. Highly antigenic peptides (VaxiJen score > 1) are shaded in grey

No.	Core Peptide	Peptide	Allele	VaxiJen Score	Position		ic50	rank
					start	end		
1	AASAHPHVT	DLPAHVAASAHPHVTV	HLA-DQA1*05:01/DQB1*03:01	0.945	75	90	18.5	0.86
	AASAHPHVT	PAHVAASAHPHVTVT	HLA-DQA1*05:01/DQB1*03:01	0.9836	74	89	22	0.62
	AASAHPHVT	AHVAASAHPHVTVT	HLA-DQA1*05:01/DQB1*03:01	0.9887	76	91	30.8	0.59
	AASAHPHVT	LPAHVAASAHPHVTVT	HLA-DQA1*05:01/DQB1*03:01	0.9952	76	90	17.4	0.74
	AASAHPHVT	PAHVAASAHPHVTV	HLA-DQA1*05:01/DQB1*03:01	0.998	75	89	28.1	0.48
	AASAHPHVT	PAHVAASAHPHVTVTP	HLA-DQA1*05:01/DQB1*03:01	1.0034	77	91	18.5	0.86
	AASAHPHVT	AHVAASAHPHVTV	HLA-DQA1*05:01/DQB1*03:01	1.0042	76	89	39.5	0.43
	AASAHPHVT	LPAHVAASAHPHVTV	HLA-DQA1*05:01/DQB1*03:01	1.0098	77	90	22.1	0.62
	AASAHPHVT	AHVAASAHPHVTVTP	HLA-DQA1*05:01/DQB1*03:01	1.0105	77	89	24.9	0.81
2	FRSAVARQQ	LVANRFRSAVARQQRLH	HLA-DRB1*08:02	0.4185	217	232	24.8	0.51
	FRSAVARQQ	LVANRFRSAVARQQRLH	HLA-DRB1*04:01	0.4185	216	232	30.5	0.28
	FRSAVARQQ	ANRFRSAVARQQRLH	HLA-DRB1*04:01	0.4288	221	237	17.7	0.23

3	FVDVLGPSP	QVAFVDVLGPSPSE	HLA-DRB1*04:01	0.5977	35	48	28.5	0.7
	FVDVLGPSP	VAFVDVLGPSPSE	HLA-DRB1*04:01	0.6011	35	47	32.8	0.63
	FVDVLGPSP	QVAFVDVLGPSPS	HLA-DRB1*04:01	0.7506	34	47	28.6	0.5
	FVDVLGPSP	VQVAFVDVLGPSPS	HLA-DRB1*04:01	0.9365	36	48	32.1	0.86
4	LAQLVANRF	HPGLAQLVANRFR	HLA-DRB1*12:01	0.4129	211	224	46.5	0.41
	LAQLVANRF	HPGLAQLVANRF	HLA-DRB1*12:01	0.4612	213	226	45.6	0.18
	LAQLVANRF	GTHPGLAQLVANRF	HLA-DRB1*12:01	0.6	212	225	46.6	0.61
5	RRVAVASFL	ANRGARRVAVASFLL	HLA-DPA1*02:01/DPB1*14:01	0.8462	171	188	41.9	0.08
	RRVAVASFL	EANRGARRVAVASFLL	HLA-DPA1*02:01/DPB1*14:01	0.9468	172	188	29	0.1
	RRVAVASFL	AEANRGARRVAVASFLL	HLA-DPA1*02:01/DPB1*14:01	0.9704	173	188	23.2	0.14
	RRVAVASFL	RAEANRGARRVAVASFLL	HLA-DPA1*02:01/DPB1*14:01	1.057	174	188	20.6	0.18
6	RVAVASFLL	NRGARRVAVASFLLAEGL	HLA-DPA1*02:01/DPB1*14:01	0.4292	178	193	12.4	0.07
	RVAVASFLL	GARRVAVASFLLAE	HLA-DPA1*02:01/DPB1*14:01	0.4639	174	189	37.3	0.04
	RVAVASFLL	NRGARRVAVASFLLAEG	HLA-DPA1*02:01/DPB1*14:01	0.4717	177	192	14.1	0.06
	RVAVASFLL	RGARRVAVASFLLAE	HLA-DPA1*02:01/DPB1*14:01	0.4922	176	190	23.1	0.02
	RVAVASFLL	ANRGARRVAVASFLLAEG	HLA-DPA1*02:01/DPB1*14:01	0.5035	175	189	12.7	0.07
	RVAVASFLL	NRGARRVAVASFLLAE	HLA-DPA1*02:01/DPB1*14:01	0.6088	178	192	16.7	0.03
	RVAVASFLL	ANRGARRVAVASFLLAE	HLA-DPA1*02:01/DPB1*14:01	0.6305	177	191	13.5	0.04
	RVAVASFLL	RGARRVAVASFLLA	HLA-DPA1*02:01/DPB1*14:01	0.6654	179	195	36.6	0.03
	RVAVASFLL	EANRGARRVAVASFLLAE	HLA-DPA1*02:01/DPB1*14:01	0.7336	176	189	12.2	0.06
	RVAVASFLL	NRGARRVAVASFLLA	HLA-DPA1*02:01/DPB1*14:01	0.7795	177	190	26	0.04
	RVAVASFLL	ANRGARRVAVASFLLA	HLA-DPA1*02:01/DPB1*14:01	0.7887	178	191	18.3	0.05
	RVAVASFLL	EANRGARRVAVASFLLA	HLA-DPA1*02:01/DPB1*14:01	0.887	179	194	14.8	0.06
	RVAVASFLL	AEANRGARRVAVASFLLA	HLA-DPA1*02:01/DPB1*14:01	0.9134	178	190	13.4	0.09
7	VRTDLPAHV	LSRGYHVRTDLPAHVA	HLA-DRB1*13:02	0.8191	67	81	6.9	0.68
	VRTDLPAHV	LSRGYHVRTDLPAHVAA	HLA-DRB1*13:02	0.8427	68	82	6.4	0.57
	VRTDLPAHV	YHVRTDLPAHVAASAH	HLA-DRB1*13:02	0.8668	68	81	7.3	0.78
	VRTDLPAHV	YHVRTDLPAHVAA	HLA-DRB1*13:02	0.8678	67	82	6.8	0.56
	VRTDLPAHV	YHVRTDLPAHVAASA	HLA-DRB1*13:02	0.8796	66	81	6.8	0.7
	VRTDLPAHV	LSRGYHVRTDLPAHV	HLA-DRB1*13:02	0.8903	68	83	8.1	0.95
	VRTDLPAHV	YHVRTDLPAHVAAS	HLA-DRB1*13:02	0.8934	66	82	6.5	0.66
	VRTDLPAHV	GYHVRTDLPAHVAASA	HLA-DRB1*13:02	1.0465	67	83	6.2	0.55
	VRTDLPAHV	GYHVRTDLPAHVA	HLA-DRB1*13:02	1.0623	67	80	6.5	0.51
	VRTDLPAHV	GYHVRTDLPAHVAA	HLA-DRB1*13:02	1.0734	65	81	5.9	0.55
	VRTDLPAHV	GYHVRTDLPAHVAAS	HLA-DRB1*13:02	1.0759	66	80	5.9	0.52
	VRTDLPAHV	RGYHVRTDLPAHVAASA	HLA-DRB1*13:02	1.0895	69	82	6.2	0.54
	VRTDLPAHV	RGYHVRTDLPAHVAAS	HLA-DRB1*13:02	1.1214	68	80	6	0.51
	VRTDLPAHV	RGYHVRTDLPAHVA	HLA-DRB1*13:02	1.1224	69	83	6.3	0.62
	VRTDLPAHV	RGYHVRTDLPAHVAA	HLA-DRB1*13:02	1.1251	69	81	5.9	0.52
	VRTDLPAHV	SRGYHVRTDLPAHVAAS	HLA-DRB1*13:02	1.1369	65	80	6.2	0.54
	VRTDLPAHV	SRGYHVRTDLPAHVA	HLA-DRB1*13:02	1.1404	69	84	6.5	0.64
	VRTDLPAHV	SRGYHVRTDLPAHVAA	HLA-DRB1*13:02	1.1416	66	79	6.2	0.55
	VRTDLPAHV	SRGYHVRTDLPAHV	HLA-DRB1*13:02	1.2457	65	79	7.6	0.84

MAFFT alignment [37] of the predicted amino acid sequence of the CbiX protein with protein sequences available in the GenBank Protein database indicated variation in the amino acid position 182. In most sequences, valine (V) is present at this position (Appendix I). However, our gene was predicted to encode for A182 instead of the common V182, as found in 99.5% of all reported CbiX protein sequences in NCBI Protein. This variation had been reported previously by a different group [11]. Analysis of the B and T – cells epitope positions indicated that this V182A variation is not located within B – cell epitopes but is present within T – cell epitopes with core peptides RRV**A**VASFL and RV**A**VASFLL (**bold A** indicates the position of the variation).

Analyses were conducted to evaluate the antigenicity and binding affinity of the HLA-II epitopes harboring the 182V/A variation site. The findings revealed that the substitution of A182 with V not only reduced the antigenicity of all HLA-II epitopes but also decreased their affinity to HLA-DPA1*02:01/DPB1*14:01, as evidenced by higher IC50 values (refer to Table 5).

Table 5. Effect of A182V variations to epitopes antigenicity and affinity to HLA-II

A182			V182			Position	
Epitope	IC50	VaxiJen Score	Epitope	IC50	VaxiJen Score	start	end
ANRGARRV**A**VASFLL	41.9	0.8462	ANRGARRV**V**VASFLL	83.6	0.8238	174	188
EANRGARRV**A**VASFLL	29	0.9468	EANRGARRV**V**VASFLL	59.5	0.9222	173	188
AEANRGARRV**A**VASFLL	23.2	0.9704	AEANRGARRV**V**VASFLL	47.8	0.9446	172	188
RAEANRGARRV**A**VASFLL	20.6	1.057	RAEANRGARRV**V**VASFLL	42.2	1.0305	171	188
NRGARRV**A**VASFLLAEGL	12.4	0.4292	NRGARRV**V**VASFLLAEGL	25.7	0.4027	175	192
GARRV**A**VASFLLAE	37.3	0.4639	GARRV**V**VASFLLAE	79.5	0.4449	177	190
NRGARRV**A**VASFLLAEG	14.1	0.4717	NRGARRV**V**VASFLLAEG	30.3	0.4459	175	191
RGARRV**A**VASFLLAE	23.1	0.4922	RGARRV**V**VASFLLAE	49.2	0.4697	176	190
ANRGARRV**A**VASFLLAEG	12.7	0.5035	ANRGARRV**V**VASFLLAEG	27	0.4771	174	191
NRGARRV**A**VASFLLAE	16.7	0.6088	NRGARRV**V**VASFLLAE	36.1	0.5842	175	190
ANRGARRV**A**VASFLLAE	13.5	0.6305	ANRGARRV**V**VASFLLAE	29.3	0.6047	174	190
RGARRV**A**VASFLLA	36.6	0.6654	RGARRV**V**VASFLLA	77	0.6464	176	189
EANRGARRV**A**VASFLLAE	12.2	0.7336	EANRGARRV**V**VASFLLAE	26.6	0.7071	173	190
NRGARRV**A**VASFLLA	26	0.7795	NRGARRV**V**VASFLLA	55.9	0.7571	175	189
ANRGARRV**A**VASFLLA	18.3	0.7887	ANRGARRV**V**VASFLLA	40	0.7641	174	189
EANRGARRV**A**VASFLLA	14.8	0.887	EANRGARRV**V**VASFLLA	32.5	0.8612	173	189
AEANRGARRV**A**VASFLLA	13.4	0.9134	AEANRGARRV**V**VASFLLA	29.3	0.8869	172	189

3-7- Epitope Population Coverage

Epitope population coverage analysis presented in Table 6 reveals that CbiX is recognized by a significant proportion of the world's population, with average population coverage of 86.14% ± 10.77%. Notably, 90% of the West African population recognises at least 2 epitopes in the CbiX protein, making them the most covered population (98.68%) by the CbiX protein-based vaccine. On the other hand, 90% of the South African population recognizes fewer than 1 epitope in the protein, making them the least covered population (52.06%) by this vaccine. South America exhibits the highest average number of recognised epitopes per HLA combination (3.98 epitopes/HLA combination), while South Africa has the lowest (1.38 epitopes/HLA combination.

Table 6. Population coverage of the HLA-I and HLA-II binding epitopes

Area	Population Coverage [a]	Average Hits [b]	PC90 [c]
Central Africa	86.92%	2.33	0.76
Central America	90.01%	1.59	1.00
East Africa	90.44%	2.59	1.02
East Asia	85.88%	2.86	0.71
Europe	93.04%	3.53	1.25
North Africa	87.17%	2.5	0.78
North America	96.00%	3.81	1.53
Northeast Asia	84.62%	2.26	0.65
Oceania	83.63%	1.91	0.61
South Africa	52.06%	1.38	0.21
South America	95.55%	3.98	1.42
South Asia	86.89%	2.6	0.76
Southeast Asia	74.38%	1.71	0.39
Southwest Asia	84.95%	2.44	0.66
West Africa	98.68%	3.63	2.01
West Indies	88.05%	2.76	0.84
Average	86.14%	2.62	0.91
Standard deviation	10.77%	0.77	0.44

[a] Projected population coverage; [b] Average number of epitope hits/HLA combinations recognized by the population; [c] Minimum number of epitope hits/HLA combinations recognized by 90% of the population

Classifying the coverage for each T-cell epitope in Table 7 reveals that the five epitopes with the highest population coverage are AASAHPHVT (Epitope #20, 56.02%±20.23% coverage), RVAVASFLL (Epitope #17, 25.92%±19.49% coverage), RRVAVASFL (Epitope #24, 23.11%±20.49% coverage), and AMADIAAQV (Epitope #3) and VLGPSPSEV (Epitope #19), which have equal coverage (21.96%±12.32%).

Table 7. Percentage of population capable of recognizing each corresponding T – cell epitope

Epitope	Population (%)																Population coverage for each epitope	
	Central Africa	Central America	East Africa	East Asia	Europe	North Africa	North America	NE Asia	Oceania	South Africa	South America	South Asia	SE Asia	SW Asia	West Africa	West Indies	Average	Standard Deviation
Epitope #1: AFLSRGYHR	1.14	0	2.11	13.92	4.72	3.4	7.09	4.08	3.28	4.56	16.56	8.42	2.65	5.82	1.71	3.48	5.18	4.49
Epitope #2: ALSCRPAIV	0	0	0	0.42	0	0.57	0.19	10.99	0	0.58	0	2.01	10.76	0	0.99	0	1.66	3.64
Epitope #3: AMADIAAQV	13.4	0	20.24	22.57	47	25.99	42.97	24.47	13.76	3.94	21.41	12.02	25	26.37	19.99	32.19	21.96	12.32
Epitope #4: AIADIAAQV	8.5	0	15.24	15.33	1.73	9.16	6.42	6.66	2.24	15.58	7.06	8.55	6.39	2.88	6.45	6.14	7.4	4.73
Epitope #5: AQLVANRFR	1.14	0	2.11	13.92	4.72	3.4	7.09	4.08	3.28	4.56	16.56	8.42	2.65	5.82	1.71	3.48	5.18	4.49
Epitope #6: EVLSALSCR	7.11	0	4.3	3.34	7.64	11.85	9.22	4.25	2.41	7.86	14.3	13.94	0.6	9.66	13.21	12.78	7.65	4.78
Epitope #7: GTRRPSGVA	6.09	0.014	11.78	1.74	2.94	12.5	6.21	4.21	1.07	11.01	1.99	7.08	2.06	5.27	7.4	0.89	5.23	3.91
Epitope #8: HPHTVTPAL	10.54	0	6.15	9.44	21.45	5.57	12.7	2.6	3.17	7.15	4.75	2.74	1.2	6.02	7.01	13.86	7.15	5.48
Epitope #9: HVAASAHHV	8.5	0	15	0	1.23	9.16	4.73	0.17	0.54	15.58	6.59	0.92	0.08	2.13	5.63	6.14	4.78	5.17
Epitope #10: IVVPAFLSR	3.41	0	2.65	1.22	6.35	6.56	6.75	1.12	2.41	7.86	12.84	11.23	0.51	6.58	8.37	6.02	5.24	3.82
Epitope #11: RARDLHTTR	1.14	0	2.11	13.92	4.72	3.4	7.09	4.08	3.28	4.56	16.56	8.42	2.65	5.82	1.71	3.48	5.18	4.49
Epitope #12: RFRSAVQQR	1.14	0	2.11	13.92	4.72	3.4	7.09	4.08	3.28	4.56	16.56	8.42	2.65	5.82	1.71	3.48	5.18	4.49
Epitope #13: RPAIVVPAF	20.8	0	10.51	22.01	30.73	16.02	25.2	6.29	5.37	12.25	8.43	8.62	3.76	11.28	24.53	23.56	14.34	9.08
Epitope #14: RPAIVVPAL	10.54	0	6.15	9.44	21.45	5.57	12.7	2.6	3.17	7.15	4.75	2.74	1.2	6.02	7.01	13.86	7.15	5.48
Epitope #15: RSAVARQQR	1.14	0	2.11	13.92	4.72	3.4	7.09	4.08	3.28	4.56	16.56	8.42	2.65	5.82	1.71	3.48	5.18	4.49
Epitope #16: RTAVQRARR	1.14	0	2.11	13.92	4.72	3.4	7.09	4.08	3.28	4.56	16.56	8.42	2.65	5.82	1.71	3.48	5.18	4.49
Epitope #17: RVAVASFLL	39.14	0.2079	39.48	21.64	27.61	9.72	15.13	23.21	14.61	3.37	53.12	46.07	7.55	12.89	73.46	6.98	25.92	19.49
Epitope #18: TVTPAGPCR	3.41	0	2.65	1.22	6.35	6.56	6.75	1.12	2.41	7.86	12.84	11.23	0.51	6.58	8.37	6.02	5.24	3.82
Epitope #19: VLGPSPSEV	13.4	0	20.24	22.57	47	25.99	42.97	24.47	13.76	3.94	21.41	12.02	25	26.37	19.99	32.19	21.96	12.32
Epitope #20: AASAHPHVT	44.17	0.8209	55.79	35.18	55.74	52.95	75.55	58.59	69.58	0	69.03	44.83	52.28	59.24	85.12	56.2	56.02	20.23
Epitope #21: FRSAVARQQ	1.8	0.2542	0	9.76	11.97	4.23	17.52	3.83	3.52	0	22.21	3.47	0.97	3.89	4.47	5.84	7.43	7.88
Epitope #22: FVDVLGPSP	1.52	0	0	3.21	11.57	3.61	15.47	1.92	1.87	0	2.28	2.76	0.36	3.67	2.44	4.47	3.45	4.23
Epitope #23: LAQLVANRF	7.25	0.0077	10.13	7.19	4.34	1.75	3.96	6	10.24	0	1.7	3.34	4.09	1.17	2.46	6.38	4.42	3.18
Epitope #24: RRVAVASFL	38.09	0.2079	37.68	21.2	22.93	2.28	11.94	22.13	13.45	0	52.48	41.71	7.03	6.22	71.9	0	23.11	20.49
Epitope #25: VRTDLPAHV	18.3	0.0287	12.76	12.1	5.46	11.17	8.3	3.43	1.02	5.91	3.78	3.59	2.19	4.93	12.34	8.68	7.3	4.86
Epitope set	86.92	90.01	90.44	85.88	93.04	87.17	96	84.62	83.63	52.06	95.55	86.89	74.38	84.95	98.68	88.05	86.14	10.77

4- Discussion

The Rv0259c gene, responsible for encoding the CbiX protein, was isolated by PCR amplification from an M. tuberculosis local strain. The obtained DNA band measured 744 bps and Sanger sequencing confirmed the sequence of Rv0259c, containing both the ATG start codon and the TGA termination codon. The gene, which was initially cloned into the pGEM®-TEasy vector (generating the recombinant pGEM®-TEasy-Rv0259c) was later subcloned into the pTrcHisA expression vector, creating the recombinant pTrcHisA-Rv0259c in E. coli BL-21(DE3) cells. Polyacrylamide gel analysis revealed that the protein expressed by E. coli BL-21(DE3) carrying pTrcHcisA-Rv0259c was a His-tagged fusion protein with molecular weight 26 kDa (Figure 8 Lane 5). Analysis of the predicted amino acid sequence generated from the Sanger DNA sequencing result indicated that the Rv0259c gene encodes for sirohydrochlorin cobaltochelatase (CbiX) protein.

CbiX is present in several species of Mycobacteria, including *M. tuberculosis, M. abscessus, and M. gastri*. A recently published study mentions that this protein is a putative lyase [12] involved in the cobalamin (vitamin B12) biosynthesis [11, 13]. Cobalamin plays crucial roles in cell metabolism, serving as a cofactor for enzymes and as a regulator of gene expression via riboswitches. The *M. tuberculosis* genome contains two vitamin B12-dependent riboswitches. The first one is encoded by the Rv1133c gene, regulating the *metE* gene expression of the cobalamin-independent methionine synthase, and the other one is encoded by the Rv0256c gene, affecting the *PPE2-cobQ1-cobU* operon, which is presumably involved in vitamin B12 synthesis [11]. The importance of this mechanism makes exploration of the Rv0256c gene as potential drug [14] and vaccine targets very attractive.

The putative 3D model of the recombinant CbiX protein, used for B cell epitopes exploration, was generated in SwissModel [24]. The model was constructed using the crystal structure of *Bacillus subtilis* sirohydrochlorin ferrochelatase (SirB protein) (PDB ID: 5ZT8.1)* [25] from the RCSB Protein Data Base† [26] as a template. With 18.50% amino acid sequence identity to CbiX protein, SirB was the best choice of template offered by SwissModel, with a GMQE score 0.57. The QMEANDisCo score of the model was 0.62±0.06.

The GMQE and QMEANDisCo global scores offer an overall assessment of model quality on a scale of 0 to 1, with higher values indicating higher expected quality. GMQE integrates properties from the target-template alignment and the template structure to predict the Local Distance Difference Test (lDDT) score of the resulting model, with its estimate being coverage-dependent [24], thus if the template used covers less than 50% of the target, the GMQE score would hardly reach 0.5. IDDT is a superposition-free score that evaluates local distance differences of all atoms in a model, with a single structure or an ensemble of equivalent structures [38]. Since the model covers amino acid residues 2 – 233 (from 247 amino acids), which is approximately 94% of all residues, the GMQE score is predominantly influenced by the target-template alignment. On the contrary, QMEANDisCo evaluates the model 'as is' without explicit coverage dependency, thus explaining the slightly higher score in the CbiX model. The QMEANDisCo global score represents the average per-residue QMEANDisCo score [39], which has been found to correlate well with the lDDT score [38].

Epitope exploration revealed that the CbiX protein contains several epitopes recognised by B and T cells. Some of these epitopes are highly immunogenic, as indicated by a VaxiJen score > 1 [34], signifying the protein's capacity in eliciting immune responses. Ten linear and eight discontinuous epitopes were revealed from B–cell epitope exploration. Exploration of T–cell epitopes revealed four highly immunogenic peptide sequences that are HLA–I binders, along with several more that are HLA-II binders.

Given that T-cells require recognition of a complex formed by HLA and epitope to trigger T-cell mediated immunity in specific individuals, it is crucial that these individuals express HLAs capable of recognizing the epitopes. HLA is highly polymorphic, it is often expressed differently, both in types and frequencies, in different ethnicities. Hence, it is important in vaccine design to consider the difference and universality of the HLAs capable of recognising the target antigens, otherwise the vaccine's population coverage would be ethnically biased [35]. For this reason, we evaluated the population coverage of T-cell epitopes present in the cloned protein. The population coverage method calculates the fraction of individuals expected to respond to a specific set of epitopes with known MHC restrictions‡ [35], based on HLA genotypic frequencies assuming non-linkage disequilibrium between HLA loci [40]. Our analysis indicated that CbiX protein could potentially vaccinate 86.14% ± 10.77% of the global population, with West Africa having the highest coverage (98.68%) and South Africa the lowest (52.06%).

Amongst the five core peptides with the highest population coverage, AMADIAAQV and VLGPSPSEV are both core peptides of HLA-I epitopes, whereas AASAHPHVT and RRVAVASFL are core peptides of HLA-II epitopes. RVAVASFLL, on the other hand, is recognised by both HLA-I and HLA-II. Considering both the core peptide population coverage and antigenicity scores of the epitopes, we are proposing that epitopes with AASAHPHVT and RRVAVASFL cores are pivotal in this vaccine. Not only do these two core peptides exhibit the highest population coverage, but the epitopes containing them also demonstrate high antigenicity scores (VaxiJen scores ranging from 0.945 to 1.0105 and 0.8462 to 1.057, respectively) (Table 8). The positions of these core peptides the protein CbiX are illustrated in Figure 12.

* *https://doi.org/10.2210/pdb5ZT8/pdb*

† *http://www.rcsb.org/*

‡ *http://tools.iedb.org/population/*

Table 8. Epitopes' core peptides with highest population coverages and their corresponding antigenicity (VaxiJen scores)

Core Peptide Sequence	Population Coverage	VaxiJen Score Range
AASAHPHVT	56.02%±20.23%	0.945 – 1.0105
RVAVASFLL	25.92%±19.49%	0.4292 – 0.9209
RRVAVASFL	23.11%±20.49%	0.8462 – 1.057
AMADIAAQV	21.96%±12.32%	0.4037
VLGPSPSEV	21.96%±12.32%	0.7381

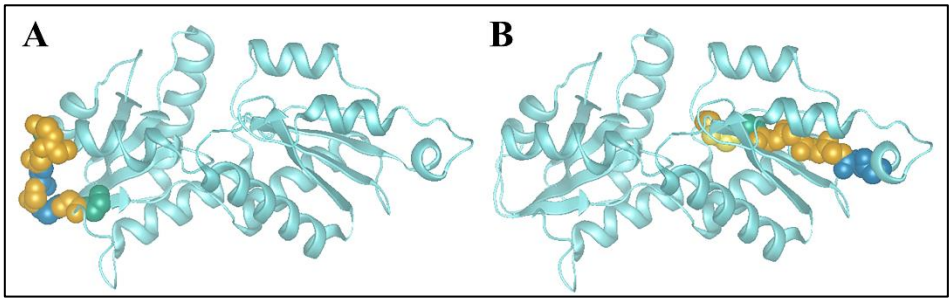

Figure 12. Positions of core peptides (in colorful beads) A. AASAHPHVT and B. RRVAVASFL

The Rv0259c gene isolated in this study encodes for CbiX protein containing A182 amino acid variation, a distinctive trait of the *M. tuberculosis* reference strain H37Rv (Accession Number: NP_214773.1). Conversely, *M. tuberculosis* CDC1551, a clinical isolate possessing cobalamin biosynthesis capability, exhibits the V182 variation, as do the majority (99.5%) of other *M. tuberculosis* strains [11]. Previous studies have indicated that consensus sequences of *M. tuberculosis* genes involved in vitamin B12 metabolism generally resemble those of *M. tuberculosis* CDC1551 [41]. Thus, it is intriguing to assume that the strain from which the Rv0259c gene was isolated is akin to *M. tuberculosis* H37Rv, a non-cobalamin-producer. The comparison of antigenicity and affinity to HLA-II between peptide epitopes containing V182 and A182 demonstrated that the A182 variation conferred higher antigenicity scores and affinity to HLA-II (refer to Table 5), suggesting that it could be a more promising vaccine candidate.

5- Conclusion

Although several studies had been published regarding its role in *M. tuberculosis* vitamin B-12 metabolism, the potentials of CbiX protein as vaccine candidate had not been widely discussed, nor explored. In this study, the Rv0259c gene was isolated from a local strain of *M. tuberculosis* using PCR methodology. Expression of the gene yielded a protein of 26 kDa on SDS-PAGE, consistent with the Sanger sequencing results, which revealed a 744bps coding sequence for the CbiX protein, encompassing the ATG start codon and the TGA termination codon. Epitope exploration targeting B and T cells, based on the predicted amino acid sequence and the putative 3D model of the protein, identified ten linear and eight discontinuous B-cell epitopes. T-cell epitope analysis unveiled four highly immunogenic peptide sequences binding to HLA-I, and even more that are HLA-II binders. *In silico* analysis considering HLA-I and HLA-II allele frequencies predicted a potential subunit vaccine using this protein could cover approximately 86.14% ± 10.77% of the global population. Two epitope cores, AASAHPHVT and RRVAVASFL, emerge as crucial elements in the protein's vaccine candidacy due to their elevated antigenicity and affinity for high-frequency HLA alleles. Although the Rv0259c gene isolated in this study contained variation resembling the minority in *M. tuberculosis* population, the 182A variation conferred higher antigenicity scores and affinity to HLA-II (refer to Table 5), suggesting that the CbiX protein could be a more promising vaccine candidate.

6- Declarations

6-1- Author Contributions

Conceptualization, A.A. and M.N.M.; methodology, A.A., A.K., R.B.L., R.A., and I.H.; validation, M.N.M.; formal analysis, A.N. and N.H.; investigation, A.A., R.A., and H.K.; writing—original draft preparation, A.A. and A.N.; writing—review and editing, A.A., M.N.M., R.I., and A.N.; supervision, M.N.M. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available in Appendix I.

6-3- Funding

This research received partial funding from Hasanuddin University (Grant Number 752/UN4.22/PL.02.00/2023).

6-4- Acknowledgements

We express our appreciation to Mrs. Kurniati Umrah Nur who has edited this document. Ms. Feny and Nadia in the Biochemistry Lab and HUM-RC, Hasanuddin University Hospital, Makassar, Indonesia, were also very helpful in preparing plasmid and recombinant DNA samples, as well as offering technical support.

6-5- Institutional Review Board and Ethical Approval Statement

All procedures involving human subjects adhered to ethical standards, with signed, written informed consent obtained in accordance with the Ethics Committee at Hasanuddin University Hospital, Makassar, Indonesia.

6-6- Informed Consent Statement

Not applicable.

6-7- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] WHO. (2022). Global Tuberculosis Report 2022. World Health Organisation (WHO), Geneva, Switzerland. Available online: https://www.who.int/publications/i/item/9789240061729 (accessed on July 2024).

[2] Gordon, S. V., & Parish, T. (2018). Microbe profile: Mycobacterium tuberculosis: Humanity's deadly microbial foe. Microbiology (United Kingdom), 164(4), 437–439. doi:10.1099/mic.0.000601.

[3] Sibuea, F., Hardhana, B., Widiantini, W. (2022). Indonesian Health Profile. Kementerian Kesehatan Republik Indonesia, Jakarta, Indonesia. Available online: https://www.kemkes.go.id/id/profil-kesehatan-indonesia-(accessed on July 2024) (In Indonesian).

[4] Tornheim, J. A., & Dooley, K. E. (2017). Tuberculosis Associated with HIV Infection. Microbiology Spectrum, 5(1). doi:10.1128/microbiolspec.tnmi7-0028-2016.

[5] Christof, C., Nußbaumer-Streit, B., & Gartlehner, G. (2020). WHO-Leitlinie: Prävention und Kontrolle von Tuberkulose-Infektionen. Das Gesundheitswesen, 82(11), 885–889. doi:10.1055/a-1241-4321. (In German).

[6] Kuan, R., Muskat, K., Peters, B., & Lindestam Arlehamn, C. S. (2020). Is mapping the BCG vaccine-induced immune responses the key to improving the efficacy against tuberculosis? Journal of Internal Medicine, 288(6), 651–660. doi:10.1111/joim.13191.

[7] Li, J., Zhao, A., Tang, J., Wang, G., Shi, Y., Zhan, L., & Qin, C. (2020). Tuberculosis vaccine development: from classic to clinical candidates. European Journal of Clinical Microbiology and Infectious Diseases, 39(8), 1405–1425. doi:10.1007/s10096-020-03843-6.

[8] Wang, R., Fan, X., Jiang, Y., Li, G., Li, M., Zhao, X., Luan, X., Deng, Y., Chen, Z., Liu, H., & Wan, K. (2023). Immunogenicity and efficacy analyses of EPC002, ECA006, and EPCP009 protein subunit combinations as tuberculosis vaccine candidates. Vaccine, 41(26), 3836–3846. doi:10.1016/j.vaccine.2023.04.003.

[9] Zhang, Y., Xu, J. C., Hu, Z. D., & Fan, X. Y. (2023). Advances in protein subunit vaccines against tuberculosis. Frontiers in Immunology, 14. doi:10.3389/fimmu.2023.1238586.

[10] Naidu, A., Nayak, S. S., Lulu S, S., & Sundararajan, V. (2023). Advances in computational frameworks in the fight against TB: The way forward. Frontiers in Pharmacology, 14. doi:10.3389/fphar.2023.1152915.

[11] Minias, A., Minias, P., Czubat, B., & Dziadek, J. (2018). Purifying selective pressure suggests the functionality of a vitamin B12 biosynthesis pathway in a global population of mycobacterium tuberculosis. Genome Biology and Evolution, 10(9), 2326–2337. doi:10.1093/gbe/evy153.

[12] Modlin, S. J., Elghraoui, A., Gunasekaran, D., Zlotnicki, A. M., Dillon, N. A., Dhillon, N., Kuo, N., Robinhold, C., Chan, C. K., Baughn, A. D., & Valafar, F. (2021). Structure-Aware Mycobacterium tuberculosis Functional Annotation Uncloaks Resistance, Metabolic, and Virulence Genes. MSystems, 6(6). doi:10.1128/msystems.00673-21.

[13] Kumar, A. (2008). Computational Annotation for Hypothetical Proteins of Mycobacterium Tuberculosis. Journal of Computer Science & Systems Biology, 01(01). doi:10.4172/jcsb.1000004.

[14] Machtel, P., Bąkowska-Żywicka, K., & Żywicki, M. (2016). Emerging applications of riboswitches – from antibacterial targets to molecular tools. Journal of Applied Genetics, 57(4), 531–541. doi:10.1007/s13353-016-0341-x.

[15] Ahmad, A., Agus, R., Massi, M. N., Handayani, I., & Karim, H. (2019). Cloning and characterization of Rv1980c gene encoding MPT64 Protein from Mycobacterium tuberculosis as a new candidate vaccine of tuberculosis. Journal of Physics: Conference Series, 1341(3), 032010. doi:10.1088/1742-6596/1341/3/032010.

[16] Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics, 13(134), 134. doi:10.1186/1471-2105-13-134.

[17] Ahmad, A., Agus, R., Massi, M. N., Natzir, R., Madhyastha, R., Madhyastha, H. K., & Maruyama, M. (2018). Cloning and expression of MPT83 gene from Mycobacterium tuberculosis in E. coli BL21 as vaccine candidate of tuberculosis: A preliminary study. Journal of Genetic Engineering and Biotechnology, 16(2), 335–340. doi:10.1016/j.jgeb.2018.04.001.

[18] Ahmad, A., Agus, R., Hidayah, N., Massi, M. N., Nurhasanah, A., & Karim, H. (2023). Cloning And Expression Of MPT83 Plus MPT64 Fusion Protein From Mycobacterium tuberculosis In Escherichia coli BL21 (DE3) Strain As Vaccine Candidate Of Tuberculosis. Rasayan Journal of Chemistry, 16(1), 297–306. doi:10.31788/RJC.2023.1618092.

[19] Sambrook, J. and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual. (3rd Ed.). Cold Spring Harbor Laboratory Press, New York, United States.

[20] Promega. (2015). pGEM®-T and pGEM®-T Easy Vector Systems. Promega, Madison, United States. Available online: https://worldwide.promega.com/products/pcr/pcr-cloning/pgem-t-easy-vector-systems/?catNum=A1360 (accessed on June 2024).

[21] Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symposium Series, 41(41), 95–98.

[22] Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. Journal of Computational Biology, 7(1–2), 203–214. doi:10.1089/10665270050081478.

[23] Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389–3402. doi:10.1093/nar/25.17.3389.

[24] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Research, 46(W1), W296–W303. doi:10.1093/nar/gky427.

[25] Fujishiro, T., Shimada, Y., Nakamura, R., & Ooi, M. (2019). Structure of sirohydrochlorin ferrochelatase SirB: The last of the structures of the class II chelatase family. Dalton Transactions, 48(18), 6083–6090. doi:10.1039/c8dt04727h.

[26] Berman, H. M. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235–242. doi:10.1093/nar/28.1.235.

[27] Karimah, N., Sulfianti, A., & Nurhasanah, A. (2022). A bioinformatic approach towards designing a human papillomavirus vaccine based on L1 capsid protein sequence of HPV45. Indian Journal of Biochemistry and Biophysics, 59(9), 927–935. doi:10.56042/ijbb.v59i9.62010.

[28] Ponomarenko, J., Bui, H. H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics, 9. doi:10.1186/1471-2105-9-514.

[29] Fleri, W. (2013) T Cell Epitopes - MHC Class I Binding Prediction Tools Description. Immune Epitope Database (IEDB), Durham, United Kingdom. Available online: https://help.iedb.org/hc/en-us/articles/114094151691-T-Cell-Epitopes-MHC-Class-I-Binding-Prediction-Tools-Description (accessed July 2024).

[30] Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Research, 48(W1), W449–W454. doi:10.1093/nar/gkaa379.

[31] Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., Buus, S., & Nielsen, M. (2009). NetMHCpan, a method for MHC class i binding prediction beyond humans. Immunogenetics, 61(1), 1–13. doi:10.1007/s00251-008-0341-z.

[32] Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., Sette, A., Peters, B., & Nielsen, M. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology, 154(3), 394–406. doi:10.1111/imm.12889.

[33] Fleri, W. (2013) T Cell Epitopes - MHC Class II Binding Prediction Tools Description. Immune Epitope Database (IEDB), Durham, United Kingdom. Available online: https://help.iedb.org/hc/en-us/articles/114094151731-T-Cell-Epitopes-MHC-Class-II-Binding-Prediction-Tools-Description (accessed July 2024).

[34] Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics, 8(4). doi:10.1186/1471-2105-8-4.

[35] Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics, 7(153). doi:10.1186/1471-2105-7-153.

[36] Sulfianti, A., Karimah, N., & Nurhasanah, A. (2023). In silico analysis of HLA-1 and HLA-2 recognition of a designed recombinant human papillomavirus vaccine based on L1 protein HPV subtype 45. Journal of Genetic Engineering and Biotechnology, 21(1), 167. doi:10.1186/s43141-023-00593-8.

[37] Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics, 20(4), 1160–1166. doi:10.1093/bib/bbx108.

[38] Mariani, V., Biasini, M., Barbato, A., & Schwede, T. (2013). IDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics, 29(21), 2722–2728. doi:10.1093/bioinformatics/btt473.

[39] Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., & Schwede, T. (2020). QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics, 36(6), 1765–1771. doi:10.1093/bioinformatics/btz828.

[40] Cun, Y., Li, C., Shi, L., Sun, M., Dai, S., Sun, L., Shi, L., & Yao, Y. (2021). COVID-19 coronavirus vaccine T cell epitope prediction analysis based on distributions of HLA class I loci (HLA-A, -B, -C) across global populations. Human Vaccines and Immunotherapeutics, 17(4), 1097–1108. doi:10.1080/21645515.2020.1823777.

[41] Minias, A., Gąsior, F., Brzostek, A., Jagielski, T., & Dziadek, J. (2021). Cobalamin is present in cells of non-tuberculous mycobacteria, but not in Mycobacterium tuberculosis. Scientific Reports, 11(1), 12267. doi:10.1038/s41598-021-91430-w.

Appendix I

Please click here to find the appendix I.